SPARSE REPRESENTATIONS FOR THREE-DIMENSIONAL RANGE DATA RESTORATION

By

Mona Mahmoudi

and

Guillermo Sapiro

IMA Preprint Series # 2280

(September 2009)

INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612-624-6066 Fax: 612-626-7370

URL: http://www.ima.umn.edu

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2009**	2. REPORT TYPE	3. DATES COVERED **00-00-2009 to 00-00-2009**
4. TITLE AND SUBTITLE **Sparse Representations for Three-Dimensional Range Data Restoration**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Minnesota,Institute for Mathematics and Its Applications,Minneapolis,MN,55455-0436**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT **see report**		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

SPARSE REPRESENTATIONS FOR THREE-DIMENSIONAL RANGE DATA RESTORATION

Mona Mahmoudi and Guillermo Sapiro

Department of Electrical and Computer Engineering, University of Minnesota
Minneapolis, MN 55455

ABSTRACT

Sparse representations of signals, in particular with learned dictionaries, are widely used for state-of-the-art audio, image, and video restoration. In this paper, the problem of denoising and occlusion restoration of 3D range data based on dictionary learning and sparse representations is explored. We consider the 3D surface obtained from a desktop range scanner as an image, where the value of each pixel represents the depth of a point on the 3D surface. Having this image, we apply techniques from dictionary learning and sparse representation to enhance the acquired 3D surface. These techniques use the spare decomposition of the overlapping patches in the image, over an adapted over-complete dictionary, for enhancing the data. We present experimental results of denoising 3D surfaces following this approach. We also propose an algorithm for filling the missing information regions on 3D scans and demonstrate its effectiveness. Our experimental results are on range data obtained from a low-cost structured-light range scanner.

***Index Terms*—** Sparse representation, 3D surface denoising, Occlusion restoration.

1. INTRODUCTION

Three-dimensional (3D) data is becoming ubiquitous. However models obtained from 3D scanners have imperfections. For example, the raw data obtained from a low-cost 3D range scanner is usually noisy and may have some occlusions or missing parts. Thus, there is an increasing need for methods for denoising and occlusion restoration of 3D surfaces in general and range data in particular. Recently, techniques based on dictionary learning for sparse representation have been widely used for image and video restoration [1, 2, 3]. In these methods, a dictionary is learned on the (overlapping) patches of the image, sparsely representing those patches, that is, each patch of the image can be well approximated only with a few atoms from the learned dictionary. In the works mentioned above, it has been shown that sparsely representing overlapping patches in the image with such learned dictionaries, and then combining them to reconstruct the image, results in an effective image denoising method.

In this work, we apply the framework of learned sparse representations in order to restore 3D surfaces, range data in particular. We also propose a new framework for filling missing information parts in 3D surfaces based on ideas similar to those presented in [2]. First, having a 3D surface scanned by a 3D range scanner, we convert it to an image whose pixel values represent the depth of each point corresponding to that pixel. Then, this image is denoised using the combination of the sparse representations of its fully overlapping patches based on the dictionary learned on the patches from the noisy data. In order to obtain the denoised 3D surface, we regenerate the 3D surface from the image by placing a point (x, y, z) on the 3D surface corresponding to each pixel (x, y) with intensity z in the image. We also introduce an iterative method to fill the holes of missing information in the range data by applying the same sparse representation method with reduced influence of the actual holes while estimating the representation. In the experiments we show the very good results obtained with this method for a low-cost scanner.

The remainder of this paper is organized as follows. In Section 2, the core algorithm for denoising 3D surfaces is presented. The method for filling the occlusions and missing information is introduced in Section 3, followed by the experimental results in Section 4. Finally we conclude the paper in Section 5.

2. SPARSE REPRESENTATION METHODS IN 3D RANGE DATA RESTORATION

In this section, we explain the basic algorithm we use for denoising 3D surfaces. In Section 2.1, the data collection process and the preprocessing of the data are explained. Some denoising algorithms for image processing based on learned sparse representation are reviewed in Section 2.2, along with the details on the specific sparse representation algorithms we use for 3D surface restoration.

This work is partially supported by ARO, NGA, ONR, DARPA, and NSF. We thank I. Ramirez and J. Mairal for providing the code for Lasso/LARS and OMP algorithms, F. Lecumberry for his help in data collection, and E. Gordon for building the structured-light scanner and installing it in our lab.

2.1. Data Collection and Preprocessing

In this work, we apply our restoration method to restore the range data collected from a low-cost structured-light 3D scanner [4]. Such low-cost family of devices produce relatively noisy range data, as well as regions with missing information due to occlusions or lack of light reflection. In addition, commonly used horizontal stripe patterns in the projected light add noise to the data with the shape of horizontal lines (Fig. 1, third column). In order to enhance the 3D data obtained from this scanner, we first convert the points on the shape to an image parallel to the camera matrix, each point with coordinates (x, y, z) corresponds to a pixel (x, y) in the image. We define the pixel value of the image as an affine function of the value of z, which is the distance from the camera, of the corresponding point. Having this natural image representation, we apply the restoration methods explained in the following sections for denoising or filling the missing information parts. Then, we show in Section 4 that if we convert the restored image back to 3D points, the result will be an enhanced 3D shape.

2.2. Denoising Surfaces Using Sparse Representation

In this section, we explain in detail the method we propose for denoising images obtained from 3D scans. Our work is based on the algorithms for image restoration using learned sparse representations (see [2] for example).

Assume $\mathbf{x_0}$ is the clean image reshaped in a vector of size N and $\mathbf{x}$ is the noisy version of $\mathbf{x_0}$. Having $\mathbf{x}$, we want to find the dictionary $\hat{\mathbf{D}}$ that "best" represents the patches in $\mathbf{x}$. In order to find $\hat{\mathbf{D}}$, the following optimization problem is addressed:

$$\begin{aligned} &\min \textstyle\sum_{ij} \|\mathbf{D}\alpha_{ij} - \mathbf{R}_{ij}\mathbf{x}\|_2^2, \\ &\text{subject to } \|d_l\|_2^2 = 1 (l = 1..k) \text{ and } |\alpha_{ij}|_p \le L, \end{aligned} \tag{1}$$

where L is a given constant; $p = 0, 1$ and $|\cdot|_p$ stand for the l_p norm; $\mathbf{D}$ is the dictionary being learned, with k atoms of length N; α_{ij} is the vector of size k coefficients corresponding to the patch at location $[i, j]$, indicating the weight of each atom from $\mathbf{D}$ in the reconstruction of the patch; and the binary matrix $\mathbf{R}_{ij}$ extracts the patch at location $[i, j]$ from the image. The minimization is performed over the dictionary $\mathbf{D}$ and the coding coefficients α.

Algorithm 1 summarizes the general approach used to solve this non-convex problem (the problem is convex on each variable when $p = 1$ but not on both at the same time).

In this work, we use the unconstrained l_1 penalty,

$$\|\mathbf{D}\alpha_{ij} - \mathbf{R}_{ij}\mathbf{x}\|_2^2 + \lambda \|\alpha_{ij}\|_1, \tag{2}$$

for each pair of $[i, j]$. In order to solve this optimization problem we used the LARS-Lasso algorithm [6], which is one of the most efficient algorithms in the literature for l_1 penalty problems. We update the dictionary using a variation of the "Method of Optimal Direction" (MOD) [5], which updates the dictionary based on the current coefficients to minimize the error in Equation (1). In particular, let $\mathbf{X}$ be a matrix whose columns are the patches of the image and $\mathbf{A}$ be a matrix whose columns are α_{ij}'s. The dictionary that minimizes Equation (1), with $\mathbf{A}$ fixed (and ignoring the atom normalization constraint), is $\mathbf{D} = \mathbf{X}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$. See [7] for more details on core components of the used optimization.

Algorithm 1 Image restoration based on sparse representation

Initialization: Let $\hat{\mathbf{D}} = (\hat{\mathbf{d}}_l)_{l \in 1..k}$ be some initial dictionary.
Dictionary Learning: Repeat J times or until convergence

- Sparse Coding: When $\hat{\mathbf{D}}$ is fixed, solve the optimization problem (Equation (1)) to find the coefficients α_{ij}. This problem is convex for $p = 1$ and can be addressed using LARS, LASSO, soft-thresholding, etc. For $p = 0$, Orthogonal Matching Pursuit is commonly used.
- Dictionary Update: In this step, we update the dictionary based on the error between the reconstructed patches and the originals [1, 5].

Image Restoration: In this part we average the reconstructed overlapping patches to restore the image. Such reconstructed patches are obtained by sparse coding with the learned dictionary.

In the last sparse coding step for the actual image restoration, after the dictionary has been learned, the best results where obtained when imposing $\|\alpha_{ij}\|_0 \le L$. We then applied the orthogonal variation of matching pursuit (OMP) [8] with $L = 2$. This combination of l_1 (via LARS) with MOD at the learning stage and l_0 with OMP at the restoration step has been experimentally found to be optimal for this and other image datasets we have tested with.[1]

Now that we have an algorithm for image denoising based on sparse representation, we can use it to denoise 3D surfaces. In order to find the 3D surface we can simply assign a point (x, y, z) to each foreground pixel (x, y) in the image whose intensity is z. The collection of these points makes the restored 3D shape.

3. FILLING MISSING INFORMATION

Similar to images, in scanning 3D data occlusion or missing information can occur. We now investigate methods for filling/inpainting the holes in 3D shape, assuming the location of the holes is known.[2] In [2], the problem of image inpainting is investigated using the sparse representations. Based on this work, we address this problem for 3D range data.

The main idea in order to fill holes is to disregard or reduce the effect of the hole pixels in the error component of Equation (2) when updating the dictionary and coefficients in the algorithm. In the first two steps of Algorithm 1, which are the "Dictionary Learning" stage, we remove all the patches

[1]We thank Julien Mairal for proposing this combination and very exhaustive testing supporting it.

[2]This can often be easily detected as lack of signal.

which have missing information, avoiding learning these irregular structures in the dictionary. In the last step, "Image Restoration," in the sparse coding part to find the optimum coefficients we define a new objective function:

$$\min \|\mathbf{R}_{ij}\mathbf{W} \otimes (\mathbf{D}\alpha_{ij} - \mathbf{R}_{ij}\mathbf{x})\|_2^2 : \ \ \|\alpha_{ij}\|_0 \leq L, \qquad (3)$$

where $\mathbf{W}$ is an adaptive matrix of weights corresponding to each pixel, see below. In this case, we first subtract the DC value of each patch before estimating the coefficients α_{ij}'s and add them to the estimated patch in the reconstruction step. For the patches containing holes we set the average value of the non-hole pixels as the DC of the patch. In order to denoise the image and fill the missing information (holes) we apply Algorithm 2 on the image obtained from the damaged data.

Algorithm 2 Iterative algorithm for filling holes on 3D surfaces

Initialization: Let $\mathbf{W}$ be the matrix of weights, which has value zero for the hole pixels and one for the rest.
Image Restoration: Find the coefficients that minimize Equation (3) for a given $\mathbf{x}$, and reconstruct the image based on these coefficients.
Hole Restoration: Find the coefficients that minimize Equation (3) for $\mathbf{x}$ being the restored image in the previous step, and reconstruct **only the holes** based on these coefficients, avoiding over-smoothing in the rest of the image.
Update Weights: Increase the weights of all the hole pixels by w_h (in our case $w_h = \frac{1}{2}$).
Image Restoration: Find the coefficients that minimize Equation (3) for $\mathbf{x}$ being the restored image in the previous restoration step, and reconstruct the image based on these coefficients.

4. EXPERIMENTAL RESULTS

In our experimental results, we apply the proposed range-data restoration framework to data obtained by a structured-light 3D scanner from some toys. This scanner finds the depth of each point based on the image of the object after some horizontal stripes projected on it. Because of these stripes, an additional noise in the collected data with the shape of horizontal lines is added to the shape (Fig. 1, third column). In some parts of the shape these lines are deeper and more difficult to remove. Also, since they exist in all the shapes, repetitive noise might be learned in the "Dictionary Learning" process. In these experiments, after collecting the data and converting the shape to an image, we normalized the intensity values and set the background to zero. In order to avoid distortions around the boundary of the shape, we reflected the values of the pixels close to the boundary inside the shape to the pixels outside the shape. We added some random holes as patches of size 10×10 pixels to each image to represent the occlusions. In both the "Dictionary Learning" and the "Image Restoration" steps we used all the patches of size 15×15 in the image. In Equation (2), λ was experimentally set to 0.16. The number of atoms in the dictionary was $k = 500$, which makes the dictionary over-complete. We set $J = 10$ in the "Dictionary Learning," and got the best results with $L = 2$ in Equation (3).

In Fig. 1, some examples of the scanned objects as image and 3D are presented. After applying Algorithm 2 to these shapes, the holes were filled and all the noise was removed except for some lines (caused by the scanning method) still left on the shapes. In order to reduce these residual imperfections, we applied Algorithm 2 again on the original shape but with a dictionary learned from the restored image. For most of the shapes reapplying Algorithm 2 improved the results. For further improving the results, we learned a dictionary on the restored images obtained from 12 shapes and used this global dictionary to denoise the original shapes with Algorithm 2. The effect of applying global dictionary was different on different shapes, a line was added back for the two dogs, improved results were obtained for the pig and not a significant change was observed for the other two shapes. Finally, Figure 2 shows some of the learned dictionaries for dog shape, pig shape, and all 12 shapes (the global one). Note that the horizontal lines had the most influence on the dictionary learned on the pig and the least influence on the dictionary learned on the dog which explains the behavior of the denoising of these shapes before and after applying the global dictionary.

5. CONCLUSIONS

In this paper, we introduced a new framework for the restoration of 3D range data. We applied sparse representation methods on images obtained from 3D surfaces in order to both denoise and fill the occluded parts of the shapes. In our experimental results we tested these methods on data obtained from a low-cost structured-light range scanner. Our experimental results demonstrate the effectiveness of these methods in denoising and filling the missing information of the 3D surfaces. We are currently working on the challenges of extending this work to full 3D shapes.

6. REFERENCES

[1] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Proc.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[2] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.

[3] M. Elad and M. Protter, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. Image Proc.*, vol. 18, no. 1, pp. 27–35, 2008.

[4] A. M. Bronstein M. M. Bronstein O. Rubinstein, Y. Honen and R. Kimmel, "3d color video camera," *Proc. Workshop on 3D Digital Imaging and Modeling (3DIM)*, 2009.

[5] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," *IEEE Int'l Conf. Acoustics, Speech and Signal Processing(ICASSP)*, vol. 5, pp. 2443–2446, 1999.

[6] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–451, 2004.

[7] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," *International Conference Machine Learning, Montreal, Canada*, June 2009.

[8] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, 1993.

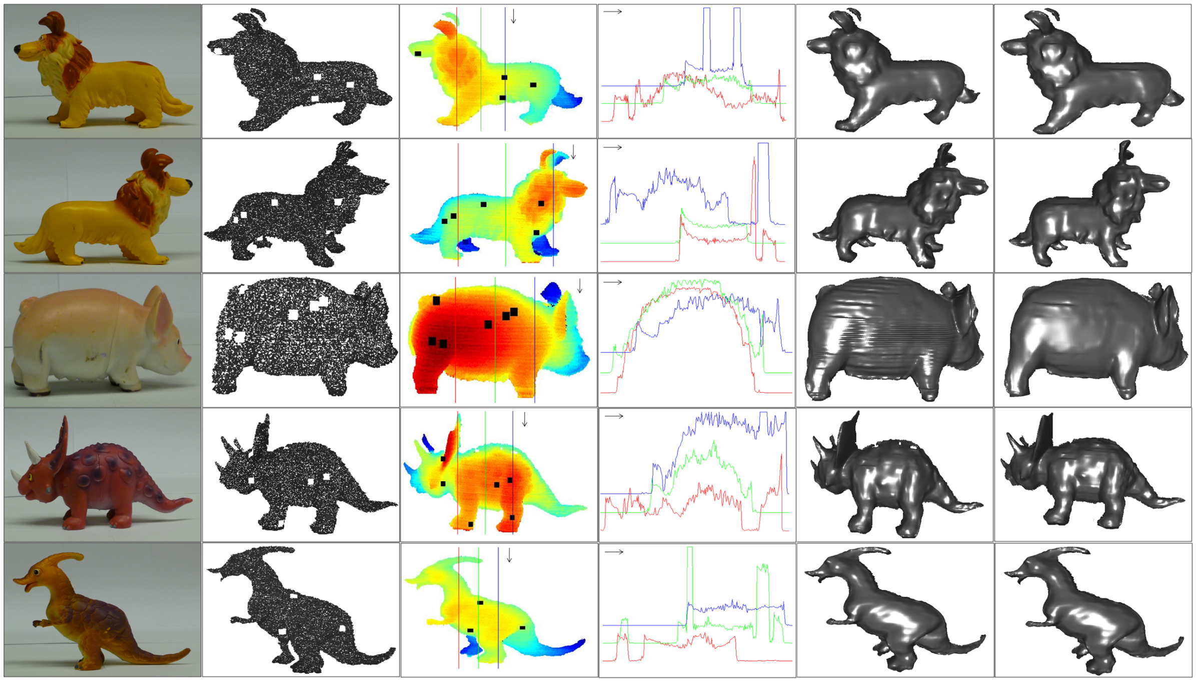

Fig. 1. *Results of the proposed method in Algorithm 2 on five shapes in the dataset. From left to right, first column shows a picture of the objects, second column shows the 3D shape obtained from the 3D scanner, and third column is the converted range-data image of the shapes in column two, see the horizontal lines. The fourth column shows the shifted intensity value (depth) of the pixels on the three lines shown in the images on the third column. The fifth column shows the restored shapes after the second run of Algorithm 2 with the dictionary learned on the restored image in the first run. The sixth column is the result of the third run of Algorithm 2 with the dictionary learned on the restored images of 12 shapes. (This is a color figure.)*

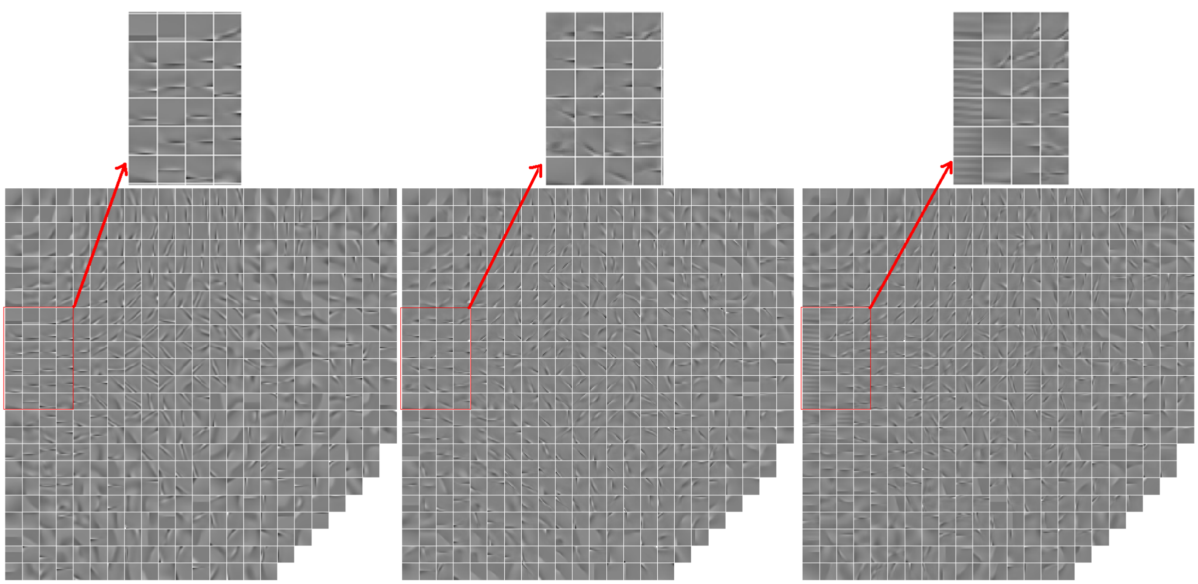

Fig. 2. *Learned range data dictionaries for 12 shapes, dog shape, and pig shape, respectively.*